

## SOLUTION TO THE LHC INVERSE PROBLEM

BRUCE KNUTESON

*MIT*

*knuteson@mit.edu*

The “LHC Inverse Problem” refers to the question of determining the underlying physical theory giving rise to the signals expected to be seen at the Large Hadron Collider. The solution to this problem (BARD) is reviewed. The combination of CDF and DØ data is motivated.

These proceedings explore the incorrectness of each of the three words in “LHC Inverse Problem.” We work backward. Sections I, II, and III discuss “Problem,” “Inverse,” and “LHC,” respectively.

### 1. Problem

The LHC Inverse Problem is fortunately no longer a problem, since it has a solution in the form of an algorithm named Bard<sup>1</sup>. BARD, named after Shakespeare, systematically weaves a set of stories to explain a particular discrepancy observed in the data.

Unfortunately no proof exists that BARD is the correct solution, the best solution, or even a good solution. The best solution to the LHC Inverse Problem will be the one leading most quickly to an underlying theory given the particular discrepancies between data and Standard Model prediction observed at the LHC. Which solution is best thus depends on how Nature actually behaves, which is what are trying to figure out.

The claim that BARD is the solution to the LHC Inverse Problem is simply the statement that BARD is the only systematic and reasonably encompassing solution to the problem so far proposed. BARD is an experimentalist’s solution to a theorist’s problem, starting directly from a specific discrepancy observed in data, and working with diagrammatic explanations that are easily visualized and understood. The essence of the algorithm is conveyed in a single sentence: Systematically draw all conceivable tree-level Feynman diagrams, introducing new particles and couplings as necessary,

where the incoming legs of the diagram are determined by the colliding particles, and the outgoing legs are determined by the events in which the signal is seen. Getting to the right answer quickly requires making reasonable assumptions to simplify the exploration of explanation space. BARD provides a convenient framework for imposing specific assumptions to reduce the size of this space.

For BARD (or any other proposed solution to the LHC Inverse Problem) to work in practice, a means of quickly, robustly, efficiently, and automatically testing any arbitrary hypothesis against the data is required. QUAERO<sup>2</sup> is the solution to this problem.

If we need BARD to tell us what we have found, then we clearly did not know what we were looking for in the first place, so how did we find it? Some means of performing a quasi-model-independent search for new physics is required. SLEUTH<sup>3,4,5</sup> or a General Search<sup>6</sup> is the solution to this problem.

Before performing a quasi-model-independent search for new physics on the high- $p_T$  tails, an understanding of the gross features of the entirety of the high- $p_T$  data is desired. VISTA<sup>7</sup> is the solution to this problem. VISTA could play an important role in the physics commissioning of the LHC experiments.

The phrase “the solution” used throughout this section unfortunately does not mean these algorithms are better than their competition, but simply that there is no competition.

## 2. Inverse

The LHC Inverse Problem is an “inverse” problem if you are used to working out the phenomenological consequences of a particular new physics scenario; inferring the new physics scenario from the phenomenological consequences is then backward. The problem is thus an inverse problem only if you are a theorist. If you are an experimentalist, it is just another problem.

The naming of this “Inverse” problem implies the forward direction involves going from a model to its phenomenological consequences. Indeed, most searches for new physics in our field start with a particular model, work out its consequences, and then fail to observe these consequences. The scientific method taught in grade school has it the other way around: make measurements on a poorly understood system, observe unexplained phenomena, and build a theory to explain existing observation and predict future observation. The grade school version is much more exciting.

### 3. LHC

The LHC Inverse Problem is an “LHC” problem because few in our field believe a surprising discovery will be made at the Fermilab Tevatron.

We have the potential for leaving a negative Tevatron legacy. It is not unlikely that after discrepancies have begun to be sorted out at the LHC it will be widely recognized by the community that a discovery could have been made at the Tevatron. This will be a tragedy and a missed opportunity. The Tevatron community should do whatever it can at small cost to ensure this does not happen.

The *de facto* discovery threshold in our field has become “ $5\sigma$ ,” with the discovery of the top quark frequently pointed to as the justification for this choice of threshold. The top quark discovery was announced when a statistical significance of  $4.8\sigma$  had been achieved at CDF<sup>8</sup> and  $4.6\sigma$  had been achieved at DØ<sup>9</sup>. From the point of view of someone outside Batavia, Illinois, the top quark was a Tevatron discovery at a statistical significance of  $4.6\sigma + 4.8\sigma = 7\sigma$ .

The Tevatron experiments are showing results on  $1 \text{ fb}^{-1}$  of data at this year’s Summer conferences. The total integrated luminosity accumulated by the Tevatron experiments, with all good run lists applied, is twice this:  $1 \text{ fb}^{-1}$  (CDF) +  $1 \text{ fb}^{-1}$  (DØ) =  $2 \text{ fb}^{-1}$ .

The Tevatron integrated luminosity is roughly doubling each year. It follows that combining CDF and DØ data will lead to a discovery one year earlier than not combining these data. With LHC turn on drawing near, there is little time to waste.

The combination of published results is not being advocated, but rather combination of the data in the exploration phase. A global analysis of CDF data should be performed, systematically searching all CDF high- $p_T$  data for discrepancies with Standard Model prediction. In parallel, a global analysis of DØ data should be performed, systematically searching all DØ high- $p_T$  data for discrepancies with Standard Model prediction. The global analyses of CDF and DØ data should be conducted within a common framework so that histograms can be filled containing both CDF and DØ events, with the overlaid Standard Model prediction the sum of the Standard Model prediction at DØ and the Standard Model prediction at CDF. VISTA and SLEUTH provide a natural framework for this effort.

Simultaneous analysis of CDF and DØ data will converge more quickly than the analysis of either alone. When a discrepancy between data and Standard Model prediction is observed in one experiment, the debugging

process must address whether the discrepancy is due to a detector effect or a poor implementation of the Standard Model prediction. On a single experiment, the debugging effort is split between these two possibilities. In a combined analysis, the presence of the effect in the other experiment focuses the debugging effort on a poor implementation of the Standard Model prediction; the absence of the effect focuses the debugging effort on a possible detector problem. A combined Tevatron analysis can be achieved within six months, if the two collaborations want this to occur.

Maximizing the chance of a Tevatron discovery requires a combined global high- $p_T$  analysis at CDF and DØ.

These proceedings have reviewed the LHC Inverse Problem. It is argued that this problem now has a solution, in the form of an algorithm named BARD; that this problem is an inverse problem only from the point of view of those working top down; and that this is a Tevatron problem until the LHC turns on. The Fermilab Tevatron will see at least three more doublings of its data: from 1 fb<sup>-1</sup> per experiment to 2 fb<sup>-1</sup>; from 2 fb<sup>-1</sup> to 4 fb<sup>-1</sup>; and a doubling as soon as a combined global analysis is achieved at CDF and DØ. Whether we can fully capitalize on this opportunity in the next two years remains to be seen.

## References

1. Bruce Knuteson and Stephen Mrenna. BARD: Interpreting new frontier energy collider physics. 2006. [hep-ph/0602101](#). 1
2. DØ Collaboration. Search for new physics using QUAERO: a general interface to DØ event data. 2001. *Phys. Rev. Lett.* **87** 231801. 2
3. DØ Collaboration. A quasi-model-independent search for new high  $p_T$  physics at DØ. 2001. *Phys. Rev. Lett.* **86** 3712. 2
4. DØ Collaboration. Search for new physics in  $e\mu X$  data at DØ using SLEUTH: A quasi model independent search strategy for new physics. 2000. *Phys. Rev.* **D 62** 092004. 2
5. DØ Collaboration. A quasi-model-independent search for new physics at large transverse momentum. 2001. *Phys. Rev.* **D 64** 012004. 2
6. H1 Collaboration. A general search for new phenomena in e p scattering at HERA. 2004. *Phys. Lett.* **B 602** (2004) 14. 2
7. Bruce Knuteson. Systematic analysis of high energy collider data. *Nucl. Instrum. Meth.* **A534** 7. 2
8. CDF Collaboration. Observation of top quark production in  $\bar{p}p$  collisions. *Phys. Rev. Lett.* **74** 2626. 3
9. DØ Collaboration. Observation of the top quark. *Phys. Rev. Lett.* **74** 2632. 3